NORTHCENTRAL UNIVERSITY
ASSIGNMENT COVER SHEET

Student: **Michael Higley-Vance**

THIS FORM MUST BE COMPLETELY FILLED IN

Follow these procedures: If requested by your instructor, please include an assignment cover sheet. This will become the first page of your assignment. In addition, your assignment header should include your last name, first initial, course code, dash, and assignment number. This should be left justified, with the page number right justified. For example:

DoeJXXX0000-1	1

Save a copy of your assignments: You may need to re-submit an assignment at your instructor's request. Make sure you save your files in accessible location.

Academic integrity: All work submitted in each course must be your own original work. This includes all assignments, exams, term papers, and other projects required by your instructor. Knowingly submitting another person's work as your own, without properly citing the source of the work, is considered plagiarism. This will result in an unsatisfactory grade for the work submitted or for the entire course. It may also result in academic dismissal from the University.

EDU7003-8 **Dr. Rebecca Watts**

Statistics **Activity #5: Correlation, Predication, Confidence, and Errors**

Comments: I feel like the only reason why I'm doing well is because I'm able to use my book and other resources! :/ I hate Chapter 8, lol. **Well, I would say that this characterizes learning. Quite often, this is how we learn. We seek out resources and we read, read, and re-read until we construct meaning from these resources. I often refer to statistics textbooks for either review or to extend my knowledge. This is not like learning your multiplication tables where you commit them to memory and for use you entire life. You retain those resources and the next time you use them, it will be much easier because you have a foundation on which to add. I hope this is helpful and encouraging!**

*****Michael, you did very well on this activity. I made some comments throughout your paper. Be sure to read them. I have highlighted them in yellow to help in locating those comments that are mine. Be sure that you locate the files that I have posted in the course discussion forum of the**

courseroom. They will be very helpful to you. Let me know if you don't locate these files and let me know if you have questions about my comments on this assignment.

Faculty Use Only

<Faculty comments here>

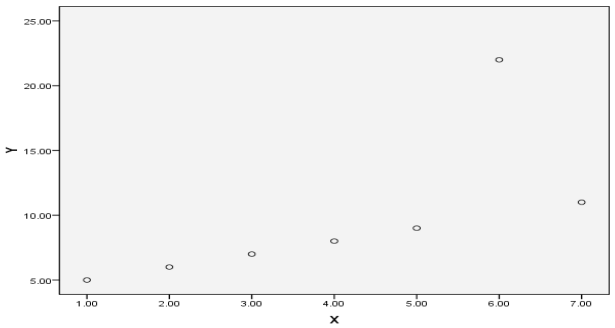
<Faculty Name> <Grade Earned> <Writing Score> <Date Graded>

Data File 4

Chapter Seven - Show all your work

Problem 1.) Look at the scatter plot below. Does it demonstrate a positive or negative correlation? Why?

Are there any outliers? What are they?

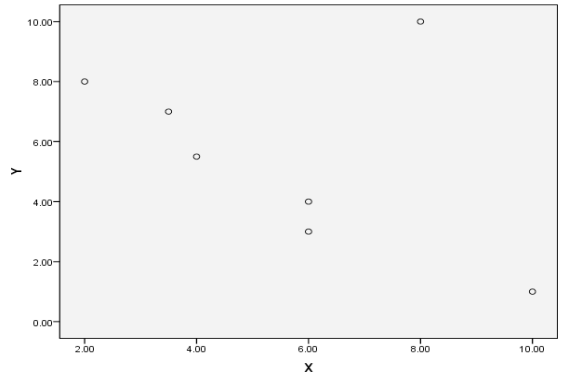


Solution: The explanatory variable is on the horizontal axis, and the response variable is on the vertical axis. There is a general upward trend indicating that as the value for the explanatory variable increases the value for the response variable also increases; therefore the scatter plot demonstrates a positive correlation with an outlier located at the point represented by the coordinate pair of 6.00 and 23.00.

12/18/13 2:17 AM
Formatted: Highlight

Problem 2.) Look at the scatter plot below. Does it demonstrate a positive or negative correlation? Why?

Are there any outliers? What are they?



Solution: Because there is a general downward trend indicating that as the value for the explanatory variable decreases the value for the response variable also decreases; therefore scatter plot demonstrates a negative correlation with an outlier of 10.00. (8.00, 10.00)

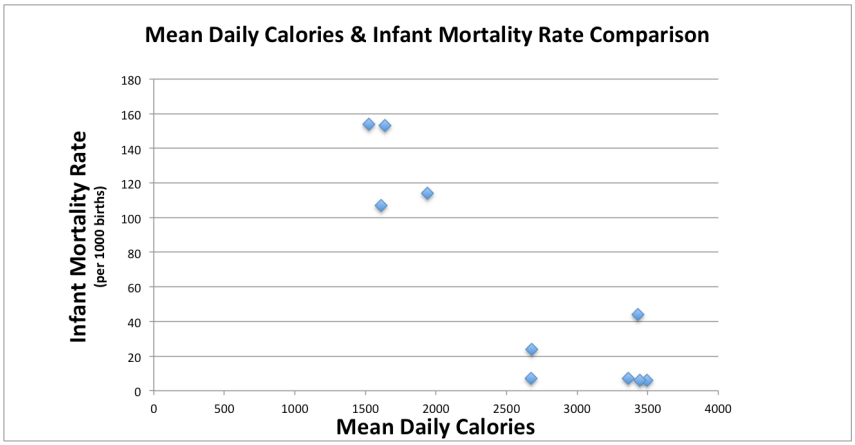
12/18/13 2:22 AM
Formatted: Highlight

Problem 3.) The following data come from your book, problem 26 on page 247.

Here is the data:

Mean daily calories	Infant Mortality Rate (per 1,000 births)
1523	154
3495	6
1941	114
2678	24
1610	107
3443	6
1640	153
3362	7
3429	44
2671	7

a) For the above data construct a scatter plot using SPSS or Excel (Follow instructions on page 244 of your textbook). What does the scatter plot show? Can you determine a type of relationship? Are there any outliers that you can see?



Solution: The scatter plot shows a negative correlation between the daily calories and the infant mortality with the outliers of 1523, 154. Yes, but I probably would not refer to this point as an outlier. If you draw a line through the center of the points, none of the points are extremely far from the line. This is how I test for outliers based on a visual review.

12/18/13 2:22 AM
Formatted: Highlight

b) Using the same data conduct a correlation analysis using SPSS or Excel. What is the correlation coefficient? Is it a strong, moderate or weak correlation? Is the correlation significant or not? If it is what does that mean?

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.901784
R Square	0.813215
Adjusted R Square	0.789867
Standard Error	382.9871
Observations	10

Solution: The correlation coefficient is .902. This is a strong correlation. In this case daily calories accounts for 81% of the variance in infant mortality, which is highly significant. Very good

Problem 4.)
Bill is doing a project for you in the marketing department. In conducting his analysis regarding consumer behavior and a new product that has come out, he tells you the correlation between these two variables is 1.09. What is your response to this analysis?

Solution: Since values for a correlation coefficient must always be between -1 and 1, it is impossible to have an $r = 1.09$. Bill's analysis appears to be inaccurate. Good!

Problem 5.)

Judy has conducted an analysis for her supervisor. The results she obtained was a correlation coefficient that was negative 0.86. Judy is confused by this number and feels that because it is negative and not positive, is means that it is bad. You are her supervisor. How would you clarify this result for Judy regarding the meaning of the correlation?

Solution: The correlation coefficient is always between -1 and 1. The strength of a correlation is determined by how close it approaches either -1 or 1. Weakness of a correlation is depicted as the value approaches 0. The sign of the coefficient (negative or positive) indicates the nature of the relationship, and does not determine if it is bad or good. Thus, an $r = .86$ is considered a very strong correlation, although it is an inverse relationship due to the negative sign of the correlation.

A positive correlation exists where high scores on one variable correspond with high scores on another. For example, the rate of obesity increases as the rate of caloric consumption increases. A negative correlation exists when high scores on one variable correspond with a low score on another. For example, performance in a basketball game decreases with the increase of anxiety. The strength and correlation range from -1.00 to +1.00. A score of -1.00 represents a correlation that is a perfect negative correlation, meaning that as a score on one variable decreases, the score on the other variable increases. A score of 0.00 represents variables that are not correlated or have no relationships, and a correlation of +1.00 represents a perfect positive correlation (the variables either increase or decrease together). Therefore, a score of -0.86 shows a negative correlation, meaning that as one variable decreases, the other variable increases.

12/18/13 2:22 AM
Formatted: Highlight

12/18/13 2:22 AM
Deleted:

12/18/13 2:22 AM
Formatted: Highlight

Problem 6.)
Explain the statement, "correlation does not imply causality."

Solution: Correlation indicates that as a measure of one variable increases or decreases a corresponding variable also increases or decreases. A correlation however, does not indicate if the first measure "causes" the second, or the second "causes" the first, or whether both variables are the outcome of a third undefined variable. Thus, correlation does not imply causality. Correct.

12/18/13 2:23 AM
Formatted: Highlight

Problem 7.)
Using the best-fit line below for prediction, answer the following questions:

a) What would you predict the price of Product X in volume of 150 to be (approximately)?

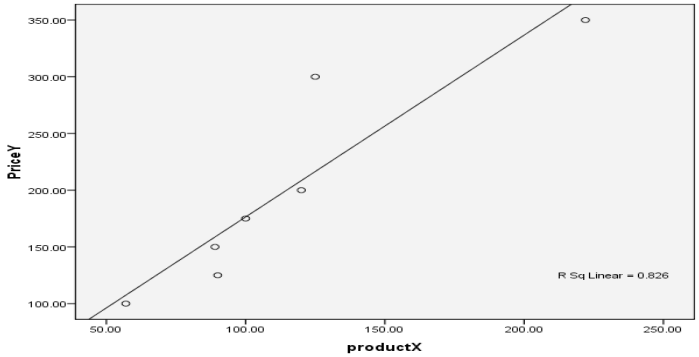
Solution A: If X = 150 then Y = 250. Predicted price = 250 okay

12/18/13 2:23 AM
Formatted: Highlight

b) What would you predict the price of Product X in volume of 100 to be (approximately)?

Solution B: If X = 100 then Y = 175. Predicted price = 175 okay

12/18/13 2:23 AM
Formatted: Highlight



Problem 8.)

You are interested in finding out if a student's ACT score is a good predictor of their final college grade point average (GPA). You have obtained the following data and are going to conduct a regression analysis. Follow instructions on page 244 of your textbook under line of best fit to conduct this analysis?

ACT	GPA
22.0	3.0
32.0	3.78
33.0	3.68
21.0	2.94
27.0	3.38
25.0	3.21
30.0	3.65

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.984275
R Square	0.968797
Adjusted R Square	0.962556
Standard Error	0.917384
Observations	7

	Coefficients	*Standard*

		Error
Intercept	-19.3364	3.746493
X Variable 1	13.76289	1.104606

In the discussion forum of the courseroom, there are files that help with these problems. For example, I provide a link to websites that help you in solving these problems. For this activity, there is a link to a website that allows you to calculate the slope and y-intercept of the regression of ACT on GPA. I pasted those results below for your review. You have the same results; however, I don't know how you are running your analyses.

12/18/13 2:28 AM
Formatted: Highlight

	X	Y
N	7	
Mean	3.3771	27.1429
Variance	0.115	22.4762
Std.Dev.	0.3391	4.7409
Std.Err.	0.1282	1.7919

r	r^2	Slope	Y Intercept	Std. Err. of Estimate
0.984	0.969	13.763	-19.3364	0.9174

t	df	P		
12.46	5		one-tailed	0.0001
			two-tailed	0.0001

0.95 and 0.99 Confidence Intervals of rho

a) What is the R? What type of relationship does it indicate (strong/weak; positive/negative)?

Solution A: There is an extremely strong positive relationship between ACT scores and GPA

b) Go to the coefficients readout. The constant is the intercept. Under that is the ACT and that is the slope. Using the straight-line formula of Y = mx + b, which you will find on page 262, you will now predict some future GPA scores: In the formula (m) is the slope; (x) is the variable that you are looking to use as a predictor; and (b) is the intercept. Predict GPA from the following ACT scores using the regression equation/straight line formula (show all your work):

1) 20
2) 25
3) 34

Solution B: For this correlation Y (ACT) = 13.76X (GPA) – 19.34

1) If X = 20; 20 = 13.76X – 19.34
   39.34 = 13.76X

X = 39.34 ÷ 13.76 = 2.86

2) If X = 25; 25 = 13.76X – 19.34
44.34 = 13.76X
X = 44.34 ÷ 13.76 = 3.22

3) If X = 34; 34 = 13.76X – 19.34
53.34 = 13.76X
X = 53.34 ÷ 13.76 = 3.88

Good work on these problems Michael!

12/18/13 2:28 AM
Formatted: Highlight

Chapter Eight - Show all your work

Problem 1.)
A sample of nine students is selected from among the students taking a particular exam. The nine students were asked how much time they had spent studying for the exam and the responses (in hours) were as follows:

18, 7, 10, 13 12, 16, 5, 20, 21

Estimate the mean study time of all students taking the exam. Round your answer to the nearest tenth of an hour, if necessary.

Solution: The mean for a sample is the average of all measures. In this sample, the mean is: (8 + 7 + 10 + 13 + 12 + 16 + 5 + 20 + 21) ÷ 9 = 122 ÷ 9 = 13.56.

Therefore, the mean study time = 13.6

Correct!

Problem 2.)
Scores on a particular test have a mean of 64.6. The distribution of sample means for samples of size 100 is normal with a mean of 64.6 and a standard deviation of 1.9. Suppose you take a sample of size 100 of test scores and find that the mean is 63. What is the z-score corresponding to this sample mean?

Solution: To calculate the z-score corresponding to this sample mean:

z = (Sample mean - population mean) / standard deviation

z = (63 - 64.6) ÷ 1.9 = - 1.6 ÷ 1.9 = - 0.84

correct

Problem 3.)
There are 349 teachers at a college. Among a sample of 110 teachers from this college, 66 have doctorates. Based on this sample, estimate the number of teachers at this college without doctorates.

Solution: The mean of the distribution of the sample proportion equals the population proportion exactly, and this distribution approaches a normal distribution as the sample increases. Since we have only been given one sample to work with, the best estimate for the population proportion p is the sample proportion:

p = (110 – 66) ÷ 110 = 44 ÷ 110 = 4 ÷ 10 = 40% don't have doctorates

For the population (N = 349) the number of teachers without doctorates would be estimated as 349 × .4 = 140

Correct

Problem 4.)
Sample size = 400; sample mean = 44; sample standard deviation = 16. What is the margin of error?

Solution: The margin of error can be calculated using the formula: $E = z \div (z \sqrt{n})$. For a 95% confidence interval the z-score is ± 1.96 (2-tailed), therefore the confidence interval would be 1.96 ÷ (1.96 √400) = 1.96 ÷ 39.2 = 0.05 or 5%

12/18/13 2:31 AM
Comment [1]: This should be s or the standard deviation. You are solving the problems correctly in that you are using the formula and substituting the correct values into the formula. You just had the wrong formula and that will cause problems. Try these again.

Margin of error = E ≈ 1.96(s/√n)

(With a margin of error for 95% confidence)
s=standard deviation of the sample 16
n=sample 400

Problem 5.)
A sample of 64 statistics students at a small college had a mean mathematics ACT score of 28 with a standard deviation of 4. Estimate the mean mathematics ACT score for all statistics students at this college. Give the 95% confidence interval.

Solution: The estimate of the population mean with a 95% confidence interval can be derived with the formula: $\mathbf{M \pm (2 \times SD \div \sqrt{n})}$
28 ± (2 × 4 ÷ √ 64) = 28 ± (8 ÷ 8) = 28 ± 1 You did this one correct.
Therefore the mean mathematics ACT score for the college exists between 27 and 29

12/18/13 2:33 AM
Formatted: Highlight

Problem 6.)
A government survey conducted to estimate the mean price of houses in a metropolitan area is designed to have a margin of error of $10,000. Pilot studies suggest that the population standard deviation is $70,000. Estimate the minimum sample size needed to estimate the population mean with the stated accuracy.

Solution: The minimum sample size needed to estimate a population mean uses this formula: $\mathbf{n \approx (z \times SD \div E)^2}$

$(1.96 \times 70{,}000 \div 10{,}000)^2 = 13.72^2 =$ 188 homes good

12/18/13 2:34 AM
Formatted: Highlight

Problem 7.)
A researcher wishes to estimate the proportion of college students who cheat on exams. A poll of 490 college students showed that 33% of them had, or intended to, cheat on examinations. Find the margin of error for the 95% confidence interval.

Solution: The margin of error from a sample is calculated using the formula:
$\mathbf{E = z \times \sqrt{((p\ (1 - p)) \div n)}}$

$1.96 \times \sqrt{((0.33 \times (1 - 0.33)) \div 490)} = 1.96 \times \sqrt{((0.33 \times 0.67) \div 490)} = 1.96 \times \sqrt{(.2211 \div 490)} = 1.96 \times \sqrt{.00045} = 1.96 \times 0.0212 = 0.0416$ or ± 4.2% good

12/18/13 2:34 AM
Formatted: Highlight

References

Bennett, J., Briggs, W., & Triola, M. (2014). *Statistical reasoning for everyday life*. (4th ed.) Boston: Pearson Education, Inc.